蜜蜂表皮蛋白 *apidermin* (*apd*)基因家族 3 个新成员的特性鉴定及昆虫 APD 家族序列特征的分析

孙亮先[1,*], 黄周英[1], 郑华军[2], 葛清秀[1], 龚丽萍[1], 陈怀宇[1]
(1. 泉州师范学院, 福建省高校分子生物学与药物化学重点实验室, 福建泉州 362000; 2. 国家人类基因组南方研究中心, 上海 201203)

摘要: Apidermin (APD)蛋白家族是一个新的昆虫结构性表皮蛋白家族。本研究结合生物信息学和 RT-PCR 扩增, 对意大利蜜蜂 *Apis mellifera ligustica*(简称"意蜂")的 *apd-1-like*, *apd-3-like* 和中华蜜蜂 *Apis cerena cerena*(简称"中蜂")的 *apd-2* 等 3 个新的 *apd* 基因的结构特征和表达进行了分析, 并分析了昆虫 APD 蛋白家族的序列特征。结果显示, 在西方蜜蜂 *Apis mellifera*(简称"西蜂")中, *apd* 基因家族的 6 个成员串联排列在基因组序列第 4 号连锁群上, 它们在 *A. m. ligustica* 雄蜂头部中的转录水平差异明显, 且其启动子序列所含顺式元件也不同。中蜂 *apd-2* 和意蜂 *apd-1-like* 都含有 3 个外显子和 2 个内含子, 而意蜂 *apd-3-like* 则由 4 个外显子和 3 个内含子组成。蛋白序列分析结果显示, 目前已知的 10 条 APD 蛋白序列 N 末端均具有相似的信号肽序列, 其成熟蛋白分子量为 6.0 ~ 37.0 kD, pI 为 6.2 ~ 10.8。其中西蜂的 APD 1 – 3、APD-like 和东方蜜蜂 *Apis cerena* 的 APD-2 等 5 条较短的多肽中疏水氨基酸残基达 52% ~ 67%, 且 Ala 含量最为丰富(占 25% ~ 34%); 而丽蝇蛹集金小蜂 *Nasonia vitripennis* 的 APD 1 – 3 和西蜂 APD-1-like, APD-3-like 等另外 5 条 APD 多肽富含 Gly(21% ~ 30%), 其序列中疏水氨基酸残基含量为 35% ~ 41%。多肽序列多重比对和系统进化分析结果显示, APD 家族可划分为 2 个亚家族。亚家族 Ⅰ 含有西蜂 APD 1 – 3 和东方蜜蜂 APD-2 等 4 条较短的多肽序列, 其 N 末端为一个长 33 aa 的保守基序; 亚家族 Ⅱ 由另外 6 条相对较长的多肽序列组成, 其 N 末端保守基序长 50 aa, C 末端保守基序长 16 aa。本文所描述的 APD 蛋白家族序列特征有助于以后从其他昆虫中鉴定新的 *apd* 基因。

关键词: 西方蜜蜂; 东方蜜蜂; 表皮蛋白; *apidermin* 基因; APD 家族; 序列分析; 生物信息学分析
中图分类号: Q966 **文献标识码**: A **文章编号**: 0454-6296(2012)01-0012-12

Characterization of three new members of the *apidermin* (*apd*) gene family from honeybees and sequence analysis of the insect APD family

SUN Liang-Xian[1,*], HUANG Zhou-Ying[1], ZHENG Hua-Jun[2], GE Qing-Xiu[1], GONG Li-Ping[1], CHEN Huai-Yu[1] (1. The Higher Educational Key Laboratory for Molecular Biology and Pharmacology of Fujian Province, Quanzhou Normal University, Quanzhou, Fujian 362000, China; 2. Chinese National Human Genome Center at Shanghai, Shanghai 201203, China)

Abstract: The apidermin (APD) protein family is a novel structural cuticular protein family of insects. To gain a better understanding of this protein family, by using bioinformatics and RT-PCR amplification, we identified three novel *apd* genes (*apd-1-like*, *apd-3-like* from *Apis mellifera ligustica*, and *apd-2* from *Apis cerana cerana*) and investigated their structural features, and then we revealed the characteristic motifs of the APD family. The results showed that a cluster of six *apd* genes were tandemly arranged on chromosome 4 of *Apis mellifera*. These *apd* genes were differentially expressed in drone head of *A. m. ligustica*, and their expression pattern was consistent with that of the *cis*-acting elements in their promoter sequences. The genomic DNA of *apd-2* from *A. c. cerana* and *apd-1-like* from *A. m. ligustica* contain three exons and two introns, while that of *apd-3-like* from *A. m. ligustica* contain four exons and three introns. As the deduced proteins of the *apd* genes were analyzed, it was found that the presently available 10 APDs possess similar N-terminal signal peptide sequences. The mature APD proteins ranged in size from 6.0 to 37.0 kD, and their pI ranged from 6.2 to 10.8. Intriguingly, five small APD proteins, including APD-2 from *A. cerana*, and APD 1 – 3 and APD-like from *A. mellifera*, were rich in hydrophobic amino acid residues (52% – 67%) with Ala being the most abundant (25% – 34%). However, the other five larger APDs, including

基金项目: 国家自然科学基金项目(31072095); 福建省高校服务海西建设重点项目(A101); 福建省教育厅省属高校项目(JA09214)
作者简介: 孙亮先, 1970 年生, 山东胶州人, 博士, 副教授, 主要从事蜜蜂功能基因组学研究
* 通讯作者 Corresponding author, E-mail: lxsunbee@163.com
收稿日期 Received: 2011-09-13; 接受日期 Accepted: 2011-11-23

APD 1 – 3 from *Nasonia vitripennis*, and APD-1-like and APD-3-like from *A. mellifera*, were Gly-rich (21% – 30%) proteins with hydrophobic amino acid residues constituting 35% – 41% of their amino acid sequences. Multiple alignment and phylogenetic analysis revealed that the APD protein family could be classified into two subfamilies. Subfamily Ⅰ, which contains four low complexity sequences (APD 1 – 3 from *A. mellifera* and APD-2 from *A. cerana*), was characterized by a 33-aa N-terminal conserved motif. The other six larger APD proteins were classified as subfamily Ⅱ, which was featured by a 50-aa N-terminal motif and a 16-aa C-terminal motif. The diagnostic features of APD protein family described here will be very helpful for identifying novel *apd* genes from arthropods.

Key words: *Apis mellifera*; *Apis cerana*; cuticular protein; *apidermin* gene; APD family; sequence analysis; bioinformatical analysis

昆虫表皮(外骨骼)是体壁上皮细胞分泌物形成的一种高度有序的层状结构，具有高等动物的皮肤和骨骼的双重功能，对运动、保持水分、维持躯体结构、抵御外界生物胁迫和非生物胁迫都有重要的作用(Willis *et al*., 2005; 刘喃喃等, 2006; 孙虹霞等, 2007)。昆虫表皮的主要成分是几丁质和多种结构性表皮蛋白(cuticular proteins, CPs)。几丁质是由 N-乙酰-D-葡糖胺单体以 β-(1,4)糖苷键聚合形成的同多糖，呈网状微丝的形式包埋在由 CPs 组成的基质内。不同表皮的几丁质分子链长和乙酰化程度差异很小，而 CPs 种类和数量的变化是影响表皮结构和机械性质的重要因素(Andersen *et al*., 1995)。对 CPs 的研究有助于阐明昆虫表皮的结构和生理功能。

Snyder 等(1982)首次报道了 4 条果蝇 *Drosophila* 的 CP 序列，至 1995 年已报道的 CP 序列为 38 条(Andersen *et al*., 1995)，至 2005 年已发表的 CP 序列已达 139 条(Willis *et al*., 2005)。近年来，随着 DNA 测序技术的快速发展，黑腹果蝇 *Drosophila melanogaster*，冈比亚按蚊 *Anopheles gambiae*，西方蜜蜂 *Apis mellifera*(简称“西蜂”)，家蚕 *Bombyx mori*，丽蝇蛹集金小蜂 *Nasonia vitripennis* 及分属不同属的 2 种蚂蚁(*Camponotus floridanus* 和 *Harpegnathos saltator*)等昆虫的全基因组序列被陆续测序，GenBank 中的昆虫基因 cDNA 和 ESTs (expressed sequence tags)序列也迅速积累。基于与已报道的表皮蛋白序列的相似性，科学家从这些核酸序列中鉴别到数以百计的表皮蛋白基因(冈比亚按蚊, 240 个; 家蚕, 221 个; 黑腹果蝇, 155 个; 丽蝇蛹集金小蜂, 76 个)(Willis, 2010; http://bioinformatics2.biol.uoa.gr/cuticleDB/)，但迄今为止从西方蜜蜂基因组中鉴别到的 CP 基因仅有 48 个(The Honeybee Genome Sequencing Consortium, 2006; Willis, 2010; 孙亮先等, 2011)，远少于从其他已测序昆虫中发现的 CP 基因数。Willis(2010)根据蛋白序列特征将昆虫结构性 CP 划分为 CPR (根据 Rebers and Riddiford Consensus 命名)、CPF (根据一段高度保守的长 44 aa 的区域命名)、CPFL (CPF like)和 *apidermin* 等 12 个家族。

昆虫表皮蛋白 apidermin(APD)家族是根据西方蜜蜂的 3 个表皮蛋白 APD 1 – 3 而命名的。目前已报道的 apidermin 家族成员蛋白有 7 个，包括西方蜜蜂的 APD 1 – 3，APD-like 和丽蝇蛹集金小蜂的 APD 1 – 3(Willis, 2010)。西方蜜蜂的 4 个 *apd* 基因串联排列在第 4 号连锁群上的一段富含 GC 区域的区域内，其成熟的产物蛋白的分子量为 6.0 ~ 12.0 kD，而丽蝇蛹集金小蜂的 3 个 *apd* 基因的成熟产物蛋白相对较大(21.6 ~ 37.0 kD)。由于 APD 蛋白序列数量有限，目前国内外尚未揭示可定义该蛋白家族的特征序列，而且尚未能在膜翅目外的其他昆虫基因组中发现类似蛋白。鉴于昆虫 CP 蛋白基因常以基因簇的形式串联排列在基因组 DNA 序列上(Togawa *et al*., 2008; Cornman *et al*., 2009)，本研究分析了西方蜜蜂已知 *apd* 基因簇附近的几个预测基因位。BLAST 相似性搜索结果揭示有 2 个预测基因的产物蛋白与丽蝇蛹集金小蜂的 APD-1 和 APD-3 序列具有较高的一致性，因此将其命名为 *apd-1-like* 和 *apd-3-like*。我们用 RT-PCR 克隆了 *apd-3-like* 基因的 cDNA 序列，并利用我们发表的一个西方蜜蜂雄蜂头部 5′LongSAGE 文库中的标签数据(Zheng *et al*., 2011)分析了西蜂 *apd* 基因家族 6 个成员的转录丰度和转录起始位点(transcription start site, TSS)。继而我们利用这 6 个西方蜜蜂 *apd* 基因的 cDNA 序列查询了中华蜜蜂 *Apis cerena cerena*(简称“中蜂”)的 ESTs 数据库(Sun *et al*., 未发表数据)和中蜂全基因组序列(东方蜜蜂 *Apis cerena* 基因组测序项目组，未发表数据)，结果在 469 162 条 ESTs 中发现了 4 条匹配 ESTs 序列，经

拼装后获得了一条编码 APD-2 的全长 cDNA，同时获得了 *apd-2* 基因的基因组 DNA 序列。在此基础上，本研究利用来源于蜜蜂和丽蝇蛹集金小蜂的 10 条 APD 多肽序列分析了 *apidermin* 蛋白家族的特征，发现该蛋白家族成员没有明显的保守序列。然而在将这 10 条蛋白序列划分为 2 个亚家族后，结果发现亚家族拥有其特征性基序。本研究首次报道了 apidermin 表皮蛋白家族的序列特征，此结果为昆虫 APD 蛋白的鉴定提供了依据。

1 材料与方法

1.1 供试材料

意大利蜜蜂 *Apis mellifera ligustica*（简称"意蜂"）和中蜂 *A. c. cerana* 蜂群饲养在泉州师范学院模式生物研究中心的实验蜂场。捕捉新出房的中蜂工蜂 1 头，意蜂工蜂和雄蜂各 1 头。撕取头胸部用于提取总 RNA。

1.2 RNA 提取与反转录合成第一链 cDNA

样品中加入 0.2 mL TRIzol 试剂（Invitrogen, San Diego, CA），充分研磨后再加入 0.8 mL TRIzol，按 Trizol 试剂盒说明书提取总 RNA。取 1 μL总 RNA 溶液于 1% 琼脂糖凝胶电泳检测。取 2 μg总 RNA，以 MMLV（Moloney-murine-leukaemia virus）逆转录酶（MBI Fermentas 公司，Vilnius, Lithuania）合成第一链 cDNA，具体方法参照 RevertAid™ Firststrand cDNA Synthesis Kit 说明书。

1.3 PCR 引物设计

以 BLAST（http://blast. ncbi. nlm. nih. gov/）对西蜂基因组序列上 *apd* 基因簇（*apd 1－3*，*apd-like*）附近的预测基因座（图 2：A）进行功能注释，根据 BLAST 结果判断预测基因座是否编码 APD 蛋白。以图 2（B）所示的 4 对 PCR 引物对西蜂预测基因座 LOC409716 的 cDNA 序列进行克隆，其中上游引物为意蜂 5′LongSAGE 文库（GEO 登录号：GSE19555）中的 4 条 5′LongSAGE 标签序列：5′-CAGTCAATCACCGGACTAC-3′，5′-AGTCAATCACCGGACTACC-3′，5′-GTCAATCACCGGACTACCA-3′和 5′-AATCACCGGACTACCAGCC-3′；下游引物根据 LOC409716 最后一个外显子设计，序列为 5′-GGGGATGATAGCTCCGTGAG-3′。西蜂预测基因座 LOC727131 的 cDNA 片段的扩增采用引物对 5′-TTTTCTCGTCTTGGTCGCA-3′和 5′-GTCCCGTGGGCAGTGATTT-3′。用引物对 5′-ATCCCTGTTAATCCTCTTCG-3′和 5′-TTCCCATCATCGTTGTTT-3′扩增中蜂 *apd*-2 基因的 cDNA 片段。引物序列委托上海基康生物技术有限公司合成。

1.4 RT-PCR 扩增与 PCR 产物克隆

取意蜂雄蜂、工蜂和中蜂工蜂的第一链 cDNA 作为模板，分别用于 RT-PCR 扩增意蜂预测基因座 LOC409716，LOC727131 和中蜂 *apd-2* 基因的 cDNA 片段。PCR 反应体系为：模板 1 μL，DreamTaq™ Green PCR Mix（2×）（Fermentas 公司）25 μL，上游和下游引物（2 pg/μL）各 1 μL，nuclease-free water 22 μL。PCR 反应程序为：95℃ 变性 3 min；95℃ 30 s，52℃ 30 s，72℃ 1.5 min；35 轮循环；72℃保温 10 min。取 2 μL PCR 产物于 1 % 琼脂糖电泳检测。将预测基因座 LOC409716 的 PCR 产物用大孔 1.5% 琼脂糖凝胶电泳后，紫外灯下切下目的条带，用 E. Z. N. A™ Gel Extraction Kit（Omega 公司）纯化回收 PCR 产物，然后用 InsTAclone™ PCR Cloning Kit（Fermentas 公司）将回收的 PCR 产物克隆进 pTZ57R/T 载体测序。具体操作参照试剂盒说明书。

1.5 西蜂 *apd* 基因家族成员转录起始位点和表达丰度分析

利用意蜂雄蜂头部 5′LongSAGE 文库（http://www. ncbi. nlm. nih. gov/geo/，编号：GSE19555）中定位在基因组序列连锁群 Contig1775 上的 5′LongSAGE 标签序列分析各 *apd* 基因的转录起始位点（TSS）和在意蜂雄蜂头部中的转录水平。标签序列的 5′末端第一个碱基在基因组上的位置可代表一个 TSS，各标签序列的拷贝数代表 RNA Pol Ⅱ 由该 TSS 起始转录的频率（Zheng *et al*., 2011）。

1.6 西蜂 *apd* 基因启动子序列的分析

在定位西蜂 *apd* 基因的 TSS 之后，用 DNASTAR 软件在基因组序列上截取优势 TSS 上游 1 kb 至下游 100 bp 的序列，在线分析（http://www. cbrc. jp/research/db/TFSEARCH. html）启动子序列含有的转录因子（TF）结合位点，并用 Photoshop 软件按比例绘制顺式作用元件分布图。

1.7 蜜蜂 *apd* 基因的内含子/外显子分析

在 NCBI 网站上以 BLAST 程序将西蜂 *apd* 基因的 cDNA 序列与西蜂基因组序列进行比对，根据 BLAST 结果在 Contig1775 上标注出各 *apd* 基因的外显子/内含子结构。同时，利用中蜂脑 ESTs 数据库中的数据（孙亮先和郑华军，未发表数据）获得中蜂 *apd-2* 基因的全长 cDNA 序列，用此 cDNA 序列 BLAST 查询中蜂全基因组序列（东方蜜蜂基因组测序项目组，未发表数据），获得中蜂 *apd-2* 基因的基

因组 DNA 序列及其外显子/内含子组成。

1.8　APD 蛋白信号肽序列预测与生化特性分析

用 SignalP 3.0 软件(Bendtsen *et al*., 2004)在线(http://www.cbs.dtu.dk/services/SignalP/)对已知 *apd* 基因的产物蛋白进行信号肽序列预测。用 DNASTAR 软件分析 APD 蛋白序列的氨基酸组成及其成熟蛋白的分子量和等电点。

1.9　APD 蛋白序列的多重比较与系统进化分析

用 MEGA 4.0 软件包中的 ClustalW 程序对 APD 蛋白序列进行多重比较，并按 Neighbor-joining（NJ）法构建系统进化树，分析时使用程序默认参数。

2　结果与分析

2.1　意蜂 *apd-3-like* 和 *apd-1-like* 以及中蜂 *apd-2* 3 个表皮蛋白基因的鉴定及实验验证

昆虫表皮蛋白基因家族常以基因簇的形式串联分布在染色体上(梁九波等, 2008)。根据此特点，本研究以 BLASTP 对西蜂基因组序列上已知 *apd* 基因簇周围的基因座进行了分析，结果发现预测基因座 LOC727131 和 LOC409716 的产物蛋白与丽蝇蛹集金小蜂的表皮蛋白 APD-1 和 APD-3 的序列相似性分别达到 51.2% 和 77.9%，因此，我们将这 2 个基因分别命名为 *apd-1-like* 和 *apd-3-like*。

为了验证预测基因座 LOC727131 和 LOC409716 的存在，本研究以 RT-PCR 对它们的 cDNA 片段进行了扩增，结果获得了预期长度的目的片段(图 1：A，B)。使用定位在 LOC409716 转录起始位点附近的 4 条 5′LongSAGE 标签序列(图 2：B)分别作为 RT-PCR 上游引物，PCR 产物经克隆测序后得到 4 条分别长 845，847，849 和 852 bp 的 cDNA 序列(GenBank 登录号分别为：JN627499，GU358200 – GU358203)。这 4 条 cDNA 序列除了 5′端相差几个碱基外，其他序列完全相同。图 1(B)为 LOC727131 的 RT-PCR 扩增片段，其长度与预期结果(561 bp)相符。

为了探索东方蜜蜂是否也存在 *apd* 基因，本研究将 6 个西蜂 *apd* 基因的 cDNA 序列用 BLAST 程序查询了中蜂 ESTs 数据库(孙亮先和郑华军，未发表数据)，结果发现了 4 条与 *apd-2* 基因相匹配的 ESTs 序列，经拼装后得到一条长 507 bp 的全长 cDNA 序列(GenBank 登录号：JN627497)。该 cDNA 编码一条长 76 aa 的多肽，其序列与西蜂 APD-2 的序列相似性为 88.3%，因此将该蛋白命名为 APD-2。图 1(C)为中蜂 *apd*-2 基因的 RT-PCR 验证结果，所得片段长度与预期结果(301 bp)吻合。用中蜂 *apd-2* 基因的全长 cDNA 序列查询中蜂全基因组序列(东方蜜蜂基因组测序项目组，未发表数据)获得了该基因的全长基因组 DNA 序列(长 1 631 bp，GenBank 登录号：JN627498)。

鉴定了 3 个 *apd* 新基因之后，本研究对来自蜜蜂的 7 个 *apd* 基因的碱基组成进行了比较(表 1)。西蜂 *apd-1-like* 和 *apd-3-like* 的 cDNA 序列中的 C + G 含量分别为 50.5% 和 64.6%，而它们的基因组 DNA 序列的 C + G 含量略低，为 42.0% 和 51.7%。东方蜜蜂 *apd-2* 基因的 cDNA 序列和基因组 DNA 序列的 GC 含量分别为 36.0% 和 45.8%。由表 1 可见，蜜蜂 *apd* 基因的基因组 DNA 序列中的 C + G 含量达 36.6%～58.6%，高于西蜂基因的平均 GC 含量（29%）（The Honeybee Genome Sequencing Consortium，2006），因此富含 GC 可能是蜜蜂 *apd* 基因的特征之一。

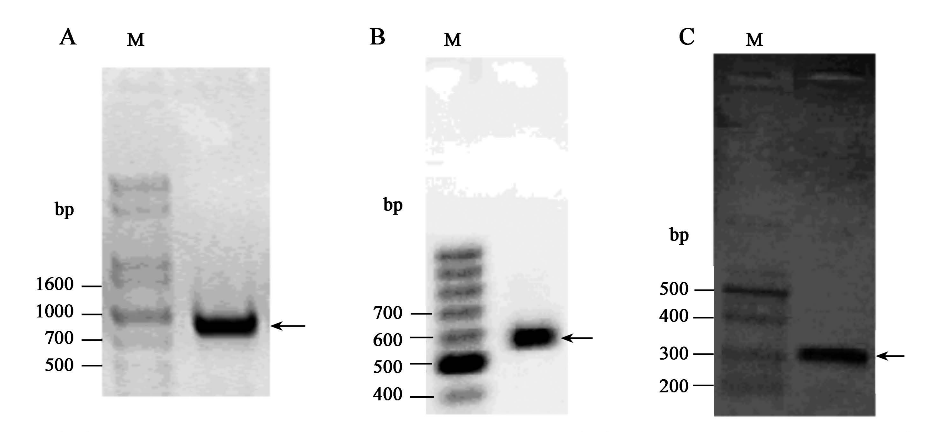


图 1　意蜂 *apd-3-like* 基因(A)和 *apd-1-like* 基因(B)以及中蜂 *apd-2* 基因(C)的 RT-PCR 结果

Fig. 1　RT-PCR results of *apd-3-like*（A）and *apd-1-like*（B）from *Apis mellifera ligustica* and *apd-2* from *Apis cerana cerana*（C）

M：DNA ladder.

表 1 蜜蜂 *apd* 基因的 cDNA 和基因组 DNA 的长度和碱基组成

Table 1 Length and base composition of cDNA and genomic DNA sequences of *apd* genes from honeybees

基因 Gene	基因的来源 Source of gene	GenBank 登录号 GenBank accession no.	A + T (%)	G + C (%)	长度(bp) Length
Apd-like	*Apis mellifera*	cDNA (NM_001174144)	30.71	69.29	534
		Genomic DNA (LOC727145)	41.43	58.57	753
Apd-1-like	*A. mellifera*	cDNA (XM_001122844)	49.52	50.48	1 153
		Genomic DNA (LOC727131)	58.00	42.00	2 038
Apd-3-like	*A. mellifera*	cDNA (JN627499)	35.39	64.61	828
		Genomic DNA (LOC409716)	48.35	51.65	3 036
Apd-1	*A. mellifera*	cDNA (XM_625244)	39.76	60.24	1 328
		Genomic DNA (LOC551367)	59.24	40.76	2 149
Apd-2	*A. mellifera*	cDNA (NM_001085346)	53.83	46.17	509
		Genomic DNA (LOC727161)	63.36	36.64	1 168
Apd-3	*A. mellifera*	cDNA (NM_001085344)	45.24	54.76	515
		Genomic DNA (LOC100037419)	57.57	42.43	839
Apd-2	*A. cerana*	cDNA (JN627497)	63.99	36.01	507
		Genomic DNA (JN627498)	54.24	45.76	1 208

2.2 西蜂 *apd* 基因的转录起始位点及在雄蜂头部中的转录水平

西蜂的 6 个 *apd* 基因串联排列在第 4 号染色体的一段长 35.9 kb 的 DNA 序列内(图 2：A)，此 DNA 区间富含 GC(41%)。根据 5′LongSAGE 标签可准确定位转录起始位点的特点(Wei *et al*., 2004; Zheng *et al*., 2011)，本研究利用雄蜂 5′LongSAGE 文库中的标签在基因组序列上定位了西蜂 *apd* 基因的 TSS，并分析了它们在意蜂雄蜂头部的转录水平(图 2：B ~ G)。5′LongSAGE 标签定位结果显示，*apd* 基因具有选择性 TSS，即 Pol Ⅱ 能在不同 TSS 上以不同的频率起始转录。如 *apd-3-like* 在雄蜂头部可由 9 个 TSS 起始转录，其中在一个优势 TSS 上起始了该基因 90% 的 mRNA(图 2：B)。有 4 条标签序列被定位在 *apd-1-like* 基因的起始密码子上游(图 2：C)，从而在转录水平上证实了 LOC727131 的存在；而 RT-PCR 结果(图 1：C)则验证了 LOC727131 的编码序列。定位在不同 *apd* 基因的 5′LongSAGE 标签拷贝数为 *apd-3-like*∶*apd*-like∶*apd-2*∶*apd-1-like*∶*apd-3*∶*apd-1* = 1 044∶429∶416∶10∶1∶0(图 2：B ~ G)，显示不同 *apd* 基因在意蜂雄蜂头部的表达丰度存在显著差异。

2.3 西蜂 *apd* 基因启动子序列的顺式作用元件

真核生物基因的启动子区一般位于 TSS 上游 300 bp 至下游 50 bp 之间。在用 5′LongSAGE 标签序列定位了西蜂 *apd* 基因的 TSS 后，本研究对其启动子序列的转录因子结合位点进行了预测，结果见图 3。由图中可见，6 个西蜂 *apd* 基因的启动子序列除含有一个典型的 TATA 框外，还都含有 3 ~ 8 个热激因子结合元件(HSE)，暗示这些基因的表达会受温度影响。*Apd-3-like* 基因的 TATA 框上游 418 bp 处有一个 CF1 结合位点，而 CF1 是类固醇受体家族的一个成员(Shea *et al*., 1990)，因此 *apd-3-like* 的表达可能会受蜕皮激素信号分子的间接调控。除 *apd-3-like* 基因外，其他 5 个蜜蜂 *apd* 基因的启动子区都含有 1 ~ 3 个 Dfd (deformed)转录因子结合位点。Dfd 是决定果蝇早期发育中形成头部体分节的一个重要转录因子(Lohmann *et al*., 2002)，因此这 5 个 *apd* 基因可能是 Dfd 蛋白的靶基因，其产物蛋白可能会在蜜蜂胚胎期头部形态形成的过程中起作用。此外，*apd-like*, *apd-2* 和 *apd-3* 等 3 个基因的启动子区含有一个 Hb (hunchback)转录因子结合位点；*apd-1-like*, *apd-3-like* 和 *apd-3* 的启动子含有 1 ~2个 Bcd (BICOID)转录因子结合位点；*apd-1-like* 的启动子还含有一个 Kr (Kruppel)结合位点，*apd-like* 的启动子含有一个同源域蛋白 Ubx (Ultrabithorax)结合位点。Hb, Kr, Bcd 和 Ubx 都是果蝇胚胎发育中重要的转录调控因子，其中 Hb 和 Bcd 在调控胚胎前端结构的形成过程中起着关键作用，而 Ubx 在第二胸节分化中起调控作用

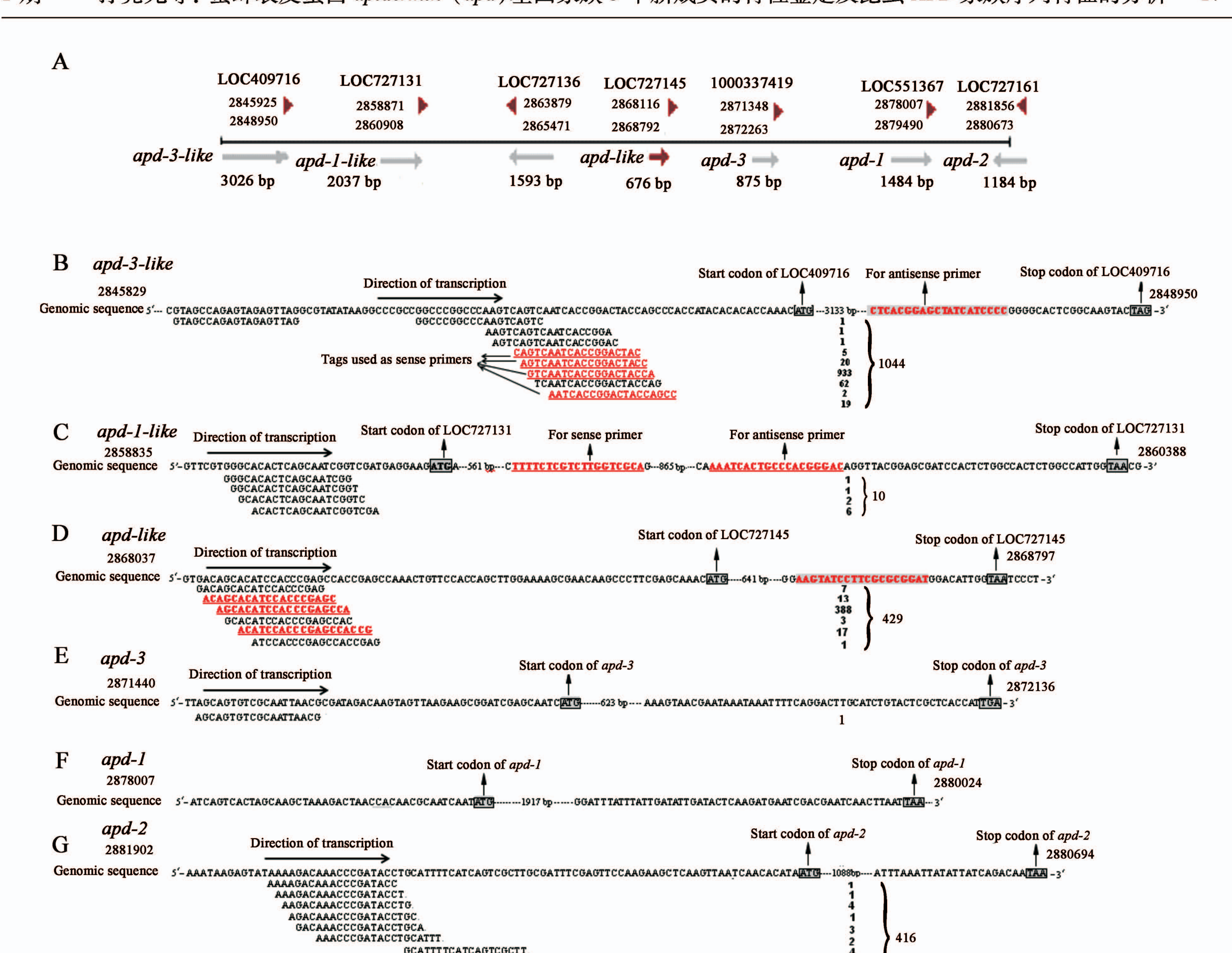


图 2　西蜂 *apd* 基因家族成员在基因组序列上的分布及其转录起始位点

Fig. 2 Distribution of the *apd* family members on genomic sequence of *Apis mellifera* and their transcription start sites

A：*apd* 基因在 contig1775 序列中的位置，长箭头表示转录方向，短箭头表示编码序列的起始位置 The relative position of *apd* genes in contig1775. Long arrows indicate the transcription direction and short arrows indicate the initiation site of start codon；B－G：定位在 *apd* 基因的转录起始位点周围的 5′LongSAGE 标签序列 The tag hits around transcription start site of *apd* genes. 标签来自于西方蜜蜂成年雄蜂头部的 5′LongSAGE 文库，标签数量代表 mRNA 由该位点起始转录的频率，编码序列的起始密码子 ATG 和终止密码子 TAA 用方框黑体标志，用作 RT-PCR 引物序列的 5′LongSAGE 标签用下划线标志。The tag sequences come from a 5′LongSAGE library of adult drone head. The number of tags indicates the initiation efficiency. The start codon and stop codon are marked with boxed-bold letter. The sequences used for designing PCR primers are underlined.

（Papatsenko and Levine，2008；Okabe-Oho *et al.*，2009；Choo *et al.*，2011）。因此蜜蜂 *apd* 基因可能会在 Hb，Bcd，Kr 和 Ubx 等转录因子的调控下参与蜜蜂胚胎形态的发生。同时，根据不同 *apd* 基因的启动子区所含有转录因子结合位点存在差异的特点，我们推测它们在蜜蜂的不同发育时期和不同组织会呈现不同表达方式。

2.4　蜜蜂 *apd* 基因的的外显子/内含子结构

图 4 比较了 7 个蜜蜂 *apd* 基因的外显子/内含子结构。从图中可见除 *apd-like* 外的另外 6 个基因的第一个外显子都很短（12～57 bp）。东方蜜蜂 *apd-2* 与西蜂 *apd 1－3* 和 *apd-1-like* 等 5 个基因都由 3 个外显子和 2 个内含子组成；而 *apd-3-like* 含 4 个外显子和 3 个内含子，且其第一个内含子长达 1 978 bp。*Apd-like* 仅含有 2 个外显子，其中第一个外显子长 389 bp。内含子边界序列分析发现，7 个蜜蜂 *apd* 基因的内含子都属于 GU-AG 型。需要说明的是，由于用于注释基因结构的 *apd-like* 和 *apd-3-like* 的 cDNA 序列 3′端为终止密码子，尚缺少 3′UTR 序列，所以所注释的最后一个外显子在 3′端尚不完整。

2.5　昆虫 APD 蛋白的生化性质

为了清晰昆虫 APD 蛋白的生化特性，本研究分析了 10 个 *apd* 基因产物蛋白的信号肽、分子量、

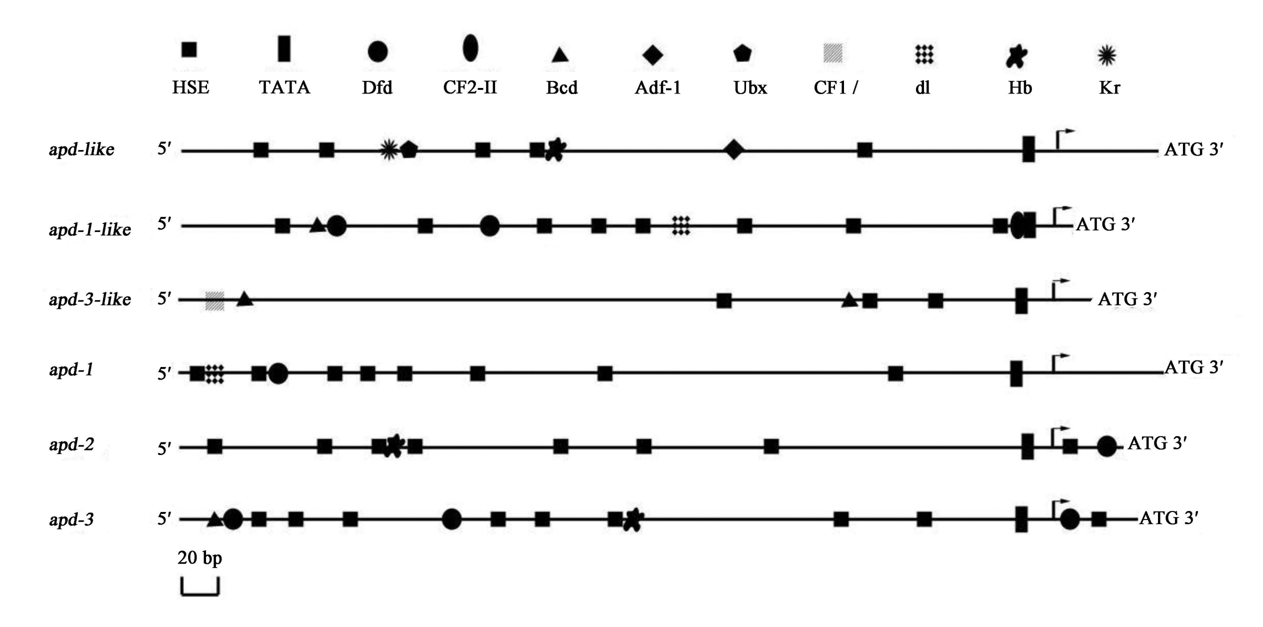


图 3 西蜂 *apd* 基因启动子序列含有的预测转录因子结合位点

Fig. 3 Scheme of putative transcription factor binding sites in promoter region of *apd* genes from *Apis mellifera*

箭头表示优势转录起始位点，ATG 表示起始密码子。The dominant transcription start site is indicated with arrow, and ATG is the start codon.

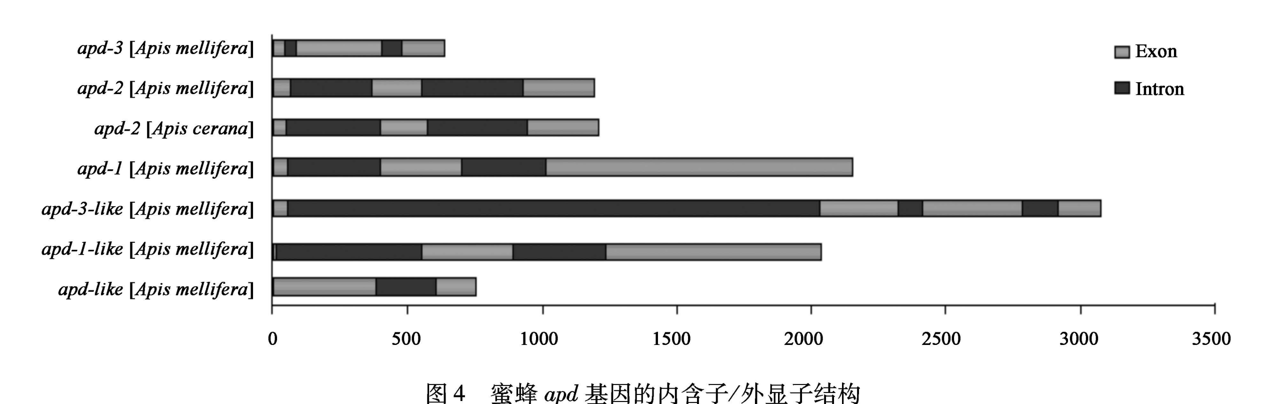


图 4 蜜蜂 *apd* 基因的内含子/外显子结构

Fig. 4 Intron/exon organization of *apd* genes from honeybees

外显子和内含子分别用灰色和黑色框表示，横坐标上的数字为实际碱基对。Exons and introns are depicted as gray and black boxes, respectively, the numbers on the *x*-axis display the actual sizes in base pairs.

等电点和 aa 组成（表 2）。信号肽预测结果显示，APD 蛋白序列 N 末端具有一段富含疏水性氨基酸的信号肽序列，表明该类 APD 蛋白为分泌型蛋白。其中，除了丽蝇蛹集金小蜂 APD-2 蛋白的信号肽序列长为 22 aa 外，其他 9 个 APD 蛋白的信号肽序列长 16 ~ 18 aa，且氨基酸顺序非常相似，因此这 9 个表皮蛋白可能会经由相似的转运机制被运输到昆虫体壁上皮细胞之外而参与外骨骼构建。切除信号肽后的成熟 APD 蛋白分子量在 6.0 ~ 37.0 kD 之间，其中东方蜜蜂 APD-2 与西蜂 APD-like、APD1 – 3 的分子量较小（6.0 ~ 12.0 kD），其未成熟多肽序列中疏水氨基酸残基占 52% ~ 67%，呈高度疏水性。西蜂 APD-1-like 和 APD-3-like 的成熟蛋白分子量分别为 18.0 kD 和 22.4 kD），与丽蝇蛹集金小蜂的 APD 1 ~ 3 的大小接近（分别为 23.1，33.7 和 21.6 kD）。这 5 个分子量相对较大的 APD 蛋白中疏水性氨基酸残基含量相对较低（36% ~ 41%）。APD 蛋白的等电点差异较大，其中西蜂的 APD 1 – 3、APD-1-like 及东方蜜蜂的 APD-2 为弱碱性蛋白，其 pI 值在 8.3 ~ 10.8之间，而其他 5 个 APD 蛋白的 pI 值在 6.2 ~ 7.1之间。由于不同 APD 蛋白的等电点和疏水性存在差异，因此在昆虫表皮内的特定 pH 环境下，这些 APD 蛋白会以不同的静电引力和疏水作用力与其他表皮蛋白或几丁质分子发生作用。

表 2　Apidermin 家族成员的生化特性

Table 2　Biochemical properties of apidermin family members

蛋白 Protein	蛋白来源 Source of protein	GenBank 登录号 GenBank accession no.	成熟蛋白分子量 Mw of mature protein	成熟蛋白 pI pI of mature protein	成熟蛋白长度(aa) Length of mature protein	信号肽长度(aa) Length of signal peptide	疏水氨基酸含量(%) Percent of hydrophobic amino acids
APD-1	*Apis mellifera*	NP_001078814.1	9 260.67	9.27	106	16	59.02
APD-2	*A. mellifera*	NP_001078815	6 110.38	10.27	61	16	66.23
APD-2	*A. cerana*	AER08513.1	6 043.25	10.83	60	16	67.11
APD-3	*A. mellifera*	NP_001078813	9 021.51	8.82	101	16	52.14
APD-like	*A. mellifera*	NP_001167615	11 951.75	7.14	134	18	54.61
APD-1-like	*A. mellifera*	XP_001122844	18 052.96	8.30	193	16	41.15
APD-3-like	*A. mellifera*	NP_001167616	22 426.60	6.17	258	17	40.00
APD-1	*Nasonia vitripennis*	NP_001166862	23 060.23	6.97	247	16	41.06
APD-2	*N. vitripennis*	XP_001605834	36 991.24	6.82	370	22	37.76
APD-3	*N. vitripennis*	XP_001606544	21 569.46	6.51	246	17	35.74

2.6　APD 蛋白序列的多序列联配和系统进化树分析

对 10 条 APD 蛋白序列的联配分析显示(图 5)，APD 蛋白序列 N 端有一段长 35 aa 的保守区：[GA]-[PVAL]-x(2)-[GA]-[PA]-[AS]-x(3)-[GVA]-[GP]-x(6)-[PA]-x(2)-[GA]-[PSG]-[AST]-x(3) -[GA]-[PSV]-x(2)-[GA]，其 C 端有一段长 18 aa 的保守性较高的区域，但 APD 家族没有明确的基序和结构域。在根据 Neighbor Joining 法构建的 APD 蛋白系统发育树中(图 6)，10 个 APD 蛋白被分为 2 组，其中西蜂 APD 1－3 和东方蜜蜂 APD-2 等 4 个分子量较小的蛋白构成系统发育树的一个分支，而丽蝇蛹集金小蜂 APD 1－3 蛋白和西蜂的其他 3 个 APD 蛋白构成发育树的另一个分支。系统进化分析结果暗示 APD 蛋白基因的祖先基因在蜜蜂和丽蝇蛹集金小蜂发生物种分化之前就已存在。

根据系统进化树分析结果，本研究将 APD 蛋白家族划分为 2 个亚家族。图 7 为这 2 个亚家族成员蛋白序列的联配分析结果。由图 7(A)可见，亚家族Ⅰ的 4 条小分子量 APD 蛋白序列具有长 33 aa 的基序：[A]-[LA]-[AV]2-x-[P]-x-[A]-x(3)-[AV]-[LV]-[AV]-x-[P]-x-[VA]-x(3)- [ILV] -[AV]-[LA]-x-[P]-x(3)-[P]-x-[LVA]-[LV]；而 APD 亚家族Ⅱ的 6 个成员蛋白序列含有 2 个保守区域，其中靠近 N 端的基序长 50 aa，其保守序列为：[G]-x(5)-[G]-[P]-x(2)-[G]-x(6)-[P]-x(2)-[G]-x(5)- [G]-x(3)- [G] -x(3)-[V]-x-[G] -x(3)- [G]-x(9)-[G]，C 端基序长 16 aa，其保守序列为：[G]-x(5)-[AV]-[VL] -[I]-x-[G]-[P] -[S]-[G]-x-[I]。

2.7　膜翅目昆虫 APD 蛋白的生物信息学鉴定

在比较蜜蜂 *adp* 基因的外显子/内含子结构(图 4)和已知的 10 条昆虫 APD 蛋白序列(图 5)时，我们发现东方蜜蜂和西蜂的 *adp-2* 在基因结构和产物蛋白序列上都非常相似，说明 *adp-2* 基因在东方蜜蜂和西蜂发生物种分化之前已经存在，同时也暗示其他蜂类昆虫也可能拥有 *adp* 基因家族。因此，我们用西蜂 *adp* 基因的 cDNA 序列查询了小蜜蜂 *Apis florea*、欧洲熊蜂 *Bombus terrestris*、美洲熊蜂 *Bombus impatiens*（http://hymenopteragenome.org/beebase/；Munoz-Torres *et al.*, 2011）和东方蜜蜂的全基因组序列，结果发现 6 个西蜂 *adp* 基因在东方蜜蜂和小蜜蜂中都高度保守；而熊蜂基因组序列中存在 *apd-3-like* 和 *apd-1* 的对应基因(BLASTN 分值分别为 396 和 119)。为了探索其他昆虫是否也存在 *adp* 基因，本研究用已知的 10 条 APD 蛋白序列查询了 NCBI 蛋白数据库，所得 BLASTP 结果见表 3。表 3 中的 12 个蚂蚁 APD 蛋白来自于蚂蚁的 4 个属：切叶蚁 *Acromyrmex echinatior*、佛罗里达弓背蚁 *Camponotus floridanus*、印度跳蚁 *Harpegnathos saltator* 和红火蚁 *Solenopsis invicta*，每个属各有 3 个 APD 蛋白基因。从基因座编号上可以看出，切叶蚁、佛罗里达弓背蚁和印度跳蚁的 *adp* 基因在基因组序列中都以基因簇的形式存在，而红火蚁的 3 个 APD 蛋白基因座分散在基因组的不同区域。美洲熊蜂的3个APD

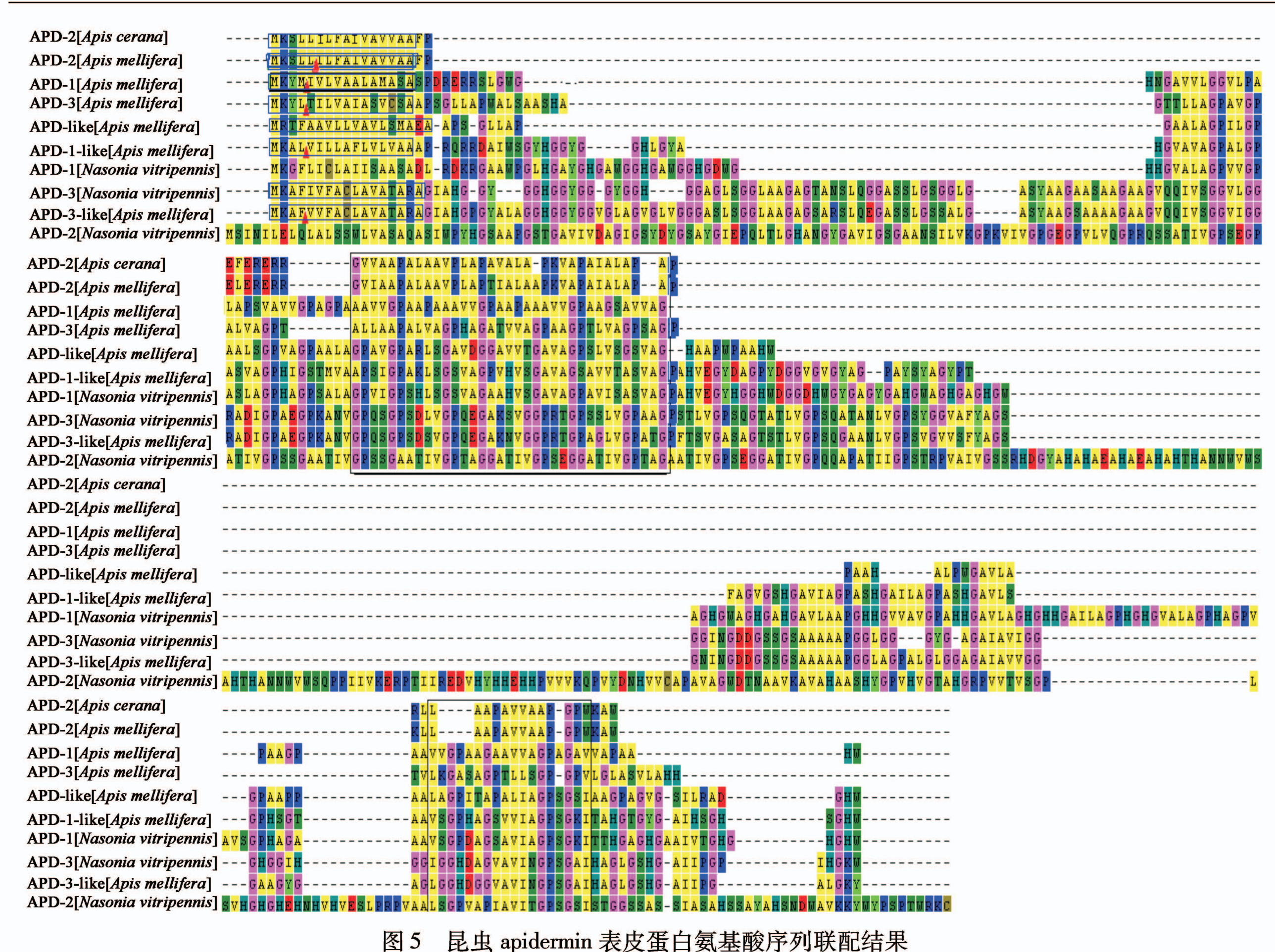


图 5 昆虫 apidermin 表皮蛋白氨基酸序列联配结果

Fig. 5 Alignment of amino acid sequences of insect apidermins

N 末端的信号肽序列和多肽序列保守区用方框标志。The N-terminal signal peptide and the conserved region are indicated with a box.

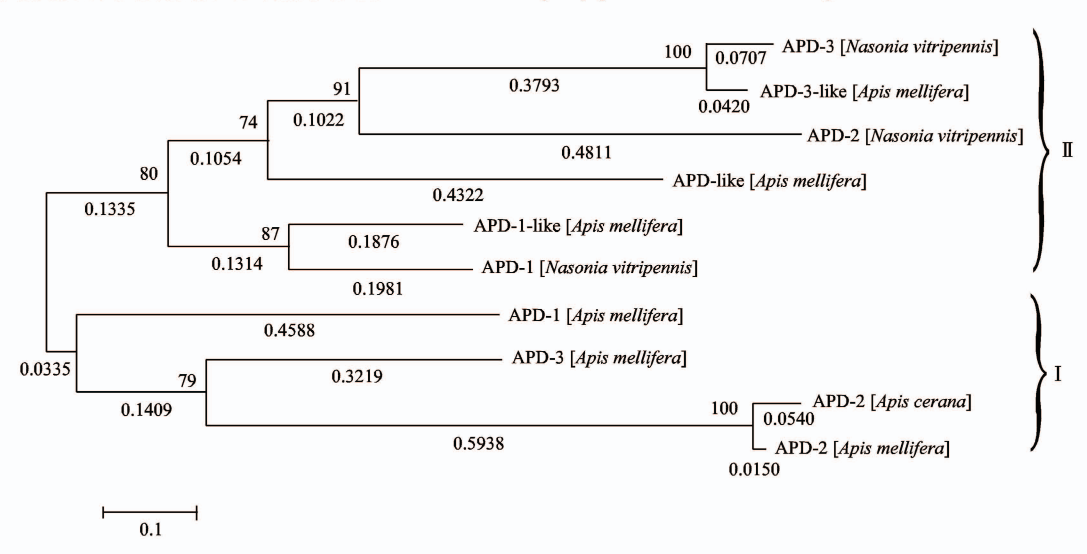


图 6 基于氨基酸序列的昆虫 apidermin 表皮蛋白系统发育树

Fig. 6 Phylogenetic tree of of insect apidermins based on their amino acid sequences

系统发育树用 MEGA 4.0 软件根据 Neighbour-Joining 法构建，枝长代表分歧度，枝上的数字为自举值 1 000 次的支持率。较短的 4 条 APD 蛋白序列被划分为组Ⅰ，另外 6 条较长的 APD 蛋白序列被划分为组Ⅱ。The tree is generated with MEGA 4.0 software using Neighbour-Joining method. The tree is drawn with branch length corresponding to partitions and branch numbers representing as percentage of bootstrap values in 1 000 sampling replicates. Four shorter APD sequences are classified into group Ⅰ, and the other six relatively longer APD sequences are classified into group Ⅱ.

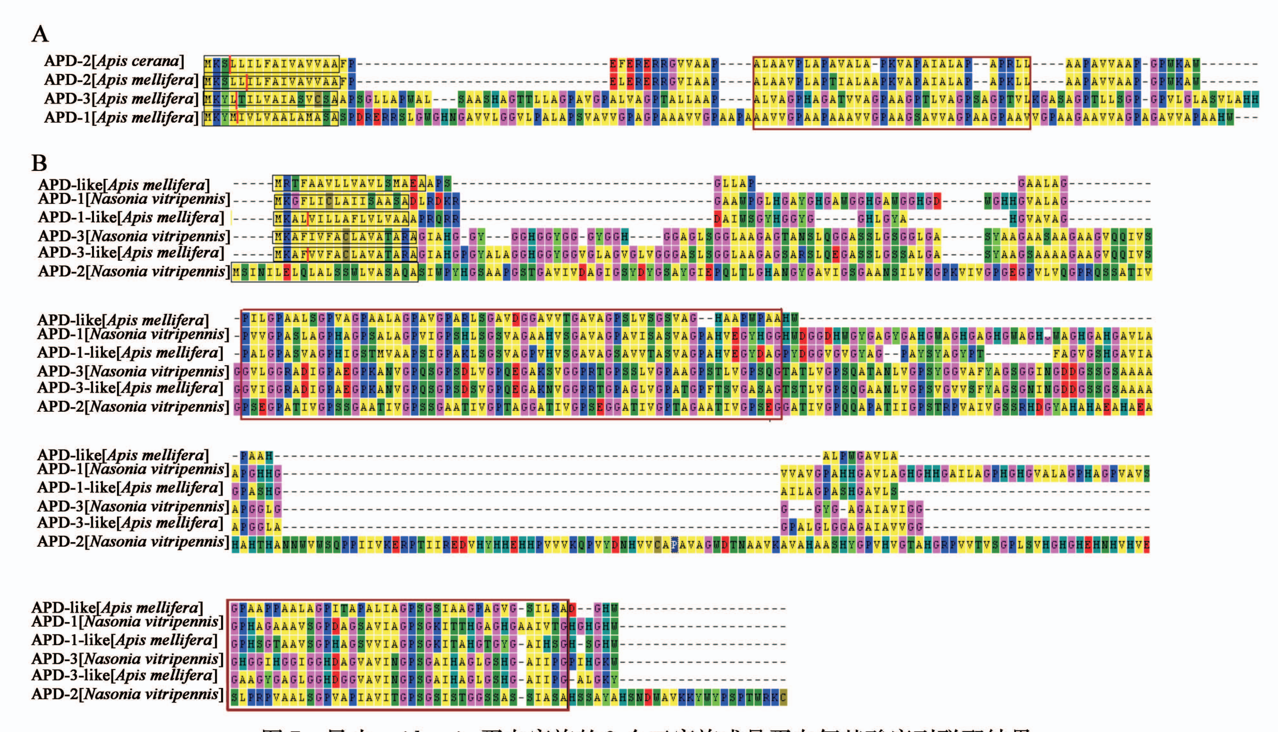


图 7　昆虫 apidermin 蛋白家族的 2 个亚家族成员蛋白氨基酸序列联配结果

Fig. 7　Alignment of amino acid sequences of two apidermin subfamilies

A：蜜蜂 APD 1 – 3 的氨基酸序列联配 Alignment of amino acid sequences of APD 1 – 3 from honeybees. B：丽蝇蛹集金小蜂 APD 1 – 3 与西方蜜蜂 APD-like，APD-1-like 和 APD-3-like 的氨基酸序列联配 Alignment of amino acid sequences of APD 1 – 3 from *Nasonia vitripennis* and APD-like，APD-1-like and APD-3-like from *Apis mellifera*. N 末端的信号肽序列和多肽序列保守区用方框标志，信号肽序列中的红色竖线表示第一个内含子在基因组序列中的位置。The N-terminal signal peptide and the conserved region are indicated with a box. The red bars in the signal peptide sequences indicate the position of the first intron in the genomic sequence.

蛋白基因座分别编号为 LOC100749732，LOC100747739 和 LOC100645440，表明前 2 个基因座在基因组序列上较为邻近；而欧洲熊蜂的 2 个 APD 蛋白基因座（LOC100651155 和 LOC100645188）则相距甚远。由于上述基因组序列近期才被测序，所以不能排除这些基因组序列存在拼装错误。本研究仅在蜜蜂、蚂蚁、熊蜂和丽蝇蛹集金小蜂等昆虫的基因组序列中鉴定到 APD 蛋白，因此 APD 可能是膜翅目昆虫所特有的一类表皮蛋白。

3　讨论

Apidermin 表皮蛋白家族是 Willis（2010）根据 3 个西蜂表皮蛋白 apidermin 1 – 3（Kucharski *et al.*，2007）和 3 个丽蝇蛹集金小蜂 apidermin 1 – 3 而划分的一个新的昆虫表皮蛋白家族。孙亮先等（2011）根据简单基因家族常串联排列在基因组序列上的特点在西蜂 *apd 1 – 3* 基因簇上游发现了 *apd-like* 基因（图 2：A）。在此基础上，本研究结合生物信息学分析和 RT-PCR 验证，又鉴定了西蜂 *apd-1-like* 和 *apd-3-like*（图 1：A，B；图 2：A），并对 6 个西蜂 *apd* 家族成员的基因结构、mRNA 转录水平及产物蛋白序列特征进行了全面比较。分析结果显示，西蜂的 6 个 *apd* 基因串联排列在第 4 号染色体的一段长 35.9 kb 的序列内，构成一个简单基因家族。5′LongSAGE 数据显示，这 6 个基因在成年意蜂雄蜂头部中的转录水平为 *apd-3-like*∶*apd-like*∶*apd-2*∶*apd-1-like*∶*apd-3*∶*apd-1* = 1 044∶429∶416∶10∶1∶0（图 2：B ~ G）。在用 5′LongSAGE 分析意蜂雄蜂头部表达谱时，我们发现 3 655 个有功能注释基因的平均表达丰度为 4.5 个标签拷贝，其中表皮蛋白 Chitin_bind_4 基因（GB16211，480 个标签拷贝）和 *apd-2*（GB10737，417 个标签拷贝）是转录丰度最高的 2 个已知基因（Zheng *et al.*，2011）。但由于当时 *apd-1-like* 和 *apd-3-like* 所在的基因座尚属预测基因，所以我们并未进行分析。本研究发现 *apd-3-like* 是雄蜂头部中转录水平水平最高的基因，其丰度是 *apd-2* 的 2.5 倍；而 *apd-1-like* 的表达水平与 *apd-2* 相似。转录水平分析结果说明，不同 *apd* 基因的产物蛋白在雄蜂头部外骨骼中的含量不同，其中 APD-3-like，APD-like 和 APD-2 可能是蜜蜂头部外骨骼的主要蛋白组分。

表 3　西蜂 apidermin 蛋白的 BLASTP 查询结果

Table 3　BLASTP results of apidermins from *Apis millifera*

查询的蛋白 Query protein		BLASTP 查询到的蛋白 Subject protein in BLASTP search			分值 Score	覆盖范围 (%) Query coverage
蛋白 Protein	登录号 GenBank accession no.	登录号 GenBank accession no.	昆虫 Insect	预测的基因座 Predicted gene cluster		
APD-like	NP_001167615	XP_003394948.1	*Bombus terrestris*	LOC100645440	189	100
		XP_003493263.1	*B. impatiens*	LOC100747739	160	100
		EFZ12167.1	*Solenopsis invicta*	SINV_08838	162	100
		EFN66202.1	*Camponotus floridanus*	EAG_13893	144	96
		EGI60900.1	*Acromyrmex echinatior*	G5I_10818	137	100
		EFN77443.1	*Harpegnathos saltator*	EAI_17469	186	100
APD-1-like	XP_001122844	XP_001606610.1	*Nasonia vitripennis*	LOC100122215	105	90
		XP_003395235.1	*B. terrestris*	LOC100651155	271	100
		XP_003493279.1	*B. impatiens*	LOC100749732	249	81
		EGI60899.1	*A. echinatior*	G5I_10817	157	80
		EFN77443.1	*H. saltator*	EAI_17469	203	98
		EFN66203.1	*C. floridanus*	EAG_13894	134	76
		EFZ12153.1	*S. invicta*	SINV_05145	119	96
APD-3-like	NP_001167616	XP_003394946.1	*B. terrestris*	LOC100645188	280	100
		EGI60898.1	*A. echinatior*	G5I_10816	209	98
		EFZ12179.1	*S. invicta*	SINV_16177	196	100
		EFN66204.1	*C. floridanus*	EAG_13895	182	72
		EFN77442.1	*H. saltator*	EAI_17468	150	35

不同西蜂 *apd* 基因的启动子区所含转录因子结合位点存在明显的差异(图 3)，意味着它们的表达调控方式可能有所不同。在 Kucharski 等(2007)的研究报道中，*adp-1* 仅在晚蛹期和成蜂早期的外骨骼表皮中表达，*adp-2* 在胚胎和成蜂的气管、食道和胃中表达，*adp-3* 在胚胎、蛹期腹部和成蜂的复眼外壳、气管、食道中表达。有趣的是，Kucharski 等(2007)的 Northern 结果表明 *adp-1* 会在工蜂脑组织中微弱表达。在发现了这一现象之后，我们用西蜂的 6 个 *adp* 基因的 cDNA 序列查询了中脑组织 EST 数据库(孙亮先和郑华军，未发表数据)，结果却在 469 162 条 ESTs 中发现了 4 条 *adp-2* 的 ESTs 序列。在拼装出中蜂 *adp-2* 基因的全长 cDNA 序列(GenBank 登录号：JN627497)后，我们用 RT-PCR 验证了拼装结果(图 1：C)。用此全长 cDNA 序列查询中蜂全基因组序列(东方蜜蜂基因组测序项目组，未发表数据)，结果得到了该基因的全长基因组 DNA 序列(GenBank 登录号：JN627498)。目前尚不清楚 *adp* 基因在蜜蜂脑组织中表达的生物学意义，或许外骨骼会向颅腔内延伸形成突起以固定和支持脑组织，而 APD 蛋白为此类支持物的组分。

Kucharski 等(2007)报道西蜂 APD 1－3 蛋白序列都含有保守四肽序列 AAPA/V，而 APD-1 和 APD-2 还具有一个富含精氨酸基序 RERR。Willis (2010)将 apidermin 划分为昆虫表皮蛋白家族之一，但未阐述该家族的序列特征。本研究对 10 条来自于西蜂、东方蜜蜂和丽蝇蛹集金小蜂的 APD 序列进行了联配分析，结果发现 APD 蛋白虽具有 2 个保守区域(图 5)，但缺乏明确的基序或结构域。然而按系统进化树分析结果将 APD 家族划分为 2 个亚家族(图 6)，则亚家族有其明确基序(图 7)。

据上讨论，我们提出如下推测：(1)膜翅目昆虫发生物种分化之前，其基因组序列中有一个较大的 *apd* 祖先基因；依据是蚂蚁、丽蝇蛹集金小蜂、熊蜂、小蜜蜂、东方蜜蜂的基因组中都含有一个与西蜂 *apd-3-like* 对应的基因，而其他 *apd* 基因的保守性较低；(2)物种分化后，*apd* 的祖先基因在不同基因组中通过基因重复形成基因簇，在此过程中新

成员基因的结构发生较大变化（如第一个内含子发生部分丢失），同时有些碱基发生突变，结果导致产物蛋白序列发生较大变化；（3）蜜蜂的 *apd* 基因家族在西方蜜蜂、东方蜜蜂、小蜜蜂发生物种分化前已形成；（4）在进化时间上，西蜂 *apd 1 - 3* 和 *apd-like* 比 *apd-3-like* 晚形成。

参考文献（References）

Andersen SO, Hojrup P, Roepstorff P, 1995. Insect cuticular proteins. *Insect Biochem. Mol. Biol.*, 25(2): 153 - 176.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S, 2004. Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, 340(4): 783 - 795.

Choo SW, White R, Russell S, 2011. Genome-wide analysis of the binding of the hox protein ultrabithorax and the hox cofactor homothorax in *Drosophila*. *PLoS ONE*, 6(4): e14778.

Cornman RS, Willis JH, 2009. Annotation and analysis of low-complexity protein families of *Anopheles gambiae* that are associated with cuticle. *Insect Biochem. Mol. Biol.*, 18(5): 607 - 622.

Kucharski R, Maleszka J, Maleszka R, 2007. Novel cuticular proteins revealed by the honey bee genome. *Insect Biochem. Mol. Biol.*, 37(2): 128 - 134.

Liang JB, Liu BL, Zhan ZG, He NJ, 2008. Bioinformation analysis of cuticular protein genes in the silkworm, *Bombyx mori*. *Science of Sericulture*, 34(3): 405 - 416. [梁九波，刘碧朗，占智高，何宁佳，2008. 家蚕表皮蛋白基因的生物信息学分析. 蚕业科学，34(3): 405 - 416]

Liu NN, Zhu F, Xu Q, Pridgeon JW, Gao XW, 2006. Behavioral change, physiological modification, and metabolic detoxification: mechanisms of insecticide resistance. *Acta Entomologica Sinica*, 49(4): 671 - 679. [刘喃喃，朱芳，徐强，Pridgeon JW，高希武，2006. 昆虫抗药性机理：行为和生理改变及解毒代谢增强. 昆虫学报，49(4): 671 - 679]

Lohmann I, McGinnis N, Bodmer M, McGinnis W, 2002. The *Drosophila Hox* gene deformed sculpts head morphology via direct regulation of the apoptosis activator reaper. *Cell*, 110(4): 457 - 466.

Munoz-Torres MC, Reese JT, Childers CP, Bennett AK, Sundaram JP, Childs KL, Anzola JM, Milshina N, Elsik CG, 2011. Hymenoptera genome database: integrated community resources for insect species of the order Hymenoptera. *Nucl. Acids Res.*, 39 (Suppl. 1): D658 - D662.

Okabe-Oho Y, Murakami H, Oho S, Sasai M, 2009. Stable, precise, and reproducible patterning of bicoid and hunchback molecules in the early *Drosophila* embryo. *PLoS Comput. Biol.*, 5(8): e1000486.

Papatsenko D, Levine MS, 2008. Dual regulation by the Hunchback gradient in the *Drosophila* embryo. *Proc. Natl. Acad. Sci. USA*, 105(8): 2901 - 2906.

Shea MJ, King DL, Conboy MJ, Mariani BD, Kafatos FC, 1990. Proteins that bind to *Drosophila* chorion *cis*-regulatory elements: a new C_2H_2 zinc finger protein and a C_2C_2 steroid receptor-like component. *Genes Dev.*, 4: 1128 - 1140.

Snyder MP, Kimbrell DA, Hunkapiller M, Hill R, Fristrom J, Davidson N, 1982. A transposable element that splits the promoter region inactivates a *Drosophila* cuticle protein gene. *Proc. Natl. Acad. Sci. USA*, 79(23): 7430 - 7434.

Sun HX, Liu Y, Zhang GR, 2007. Effects of heavy metal pollution on insects. *Acta Entomologica Sinica*, 50(2): 178 - 185. [孙虹霞，刘颖，张古忍，2007. 重金属污染对昆虫生长发育的影响. 昆虫学报，50(2): 178 - 185]

Sun LX, Huang ZY, Zheng HJ, You YL, 2011. TSS identification and cDNA cloning of a cuticular protein gene *apd*-like from the Western honeybee (*Apis mellifera*). *Acta Entomologica Sinica*, 54(2): 224 - 231. [孙亮先，黄周英，郑华军，游燕琳，2011. 西方蜜蜂表皮蛋白基因 *apd-like* 转录起始位点的定位及 cDNA 序列的分析. 昆虫学报，54(2): 224 - 231]

The Honeybee Genome Sequencing Consortium, 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature*, 443 (7114): 931 - 949.

Togawa T, Dunn WA, Emmons AC, Nagao J, Willis JH, 2008. Developmental expression patterns of cuticular protein genes with the R&R Consensus from *Anopheles gambiae*. *Insect Biochem. Mol. Biol.*, 38(5): 508 - 519.

Wei CL, Ng P, Chiu KP, Wong CH, Ang CC, Lipovich L, Liu ET, Ruan Y, 2004. 5′ Long serial analysis of gene expression (LongSAGE) and 3′ LongSAGE for transcriptome characterization and genome annotation. *Proc. Natl. Acad. Sci. USA*, 101(32): 11701 - 11706.

Willis JH, 2010. Structural cuticular proteins from arthropods: annotation, nomenclature, and sequence characteristics in the genomics era. *Insect Biochem. Mol. Biol.*, 40(3): 189 - 204.

Willis JH, Iconomidou VA, Smith RF, Hamodrakas SJ, 2005. Cuticular proteins. In: Gilbert LI, Latrou K, Gill SS eds. Comprehensive Molecular Insect Science. Vol. 4. Elsevier, Oxford. 79 - 109.

Zheng H, Sun L, Peng W, Shen Y, Wang Y, Xu B, Gu W, Chen S, Huang Z, Wang S, 2011. Global identification of transcription start sites in the genome of *Apis mellifera* using 5′LongSAGE. *J. Exp. Zool.* (*Mol. Dev. Evol.*), 316B(7): 500 - 514.

（责任编辑：赵利辉）